CHAPTER IV
Results

4.1 Total RNA extraction

Young leaves of *C. ternatea* L. and calli of *A. lakoocha* Rox. were used for RNA sources. Total RNAs were extracted by RNeasy Plant Mini Kit (Qiagen). The quality and concentration of all total RNA samples were determined on agarose gel as shown in Figure 14, and UV absorption at wavelength of 260 nm, respectively. The concentration of the total RNA was 320 and 225 $\mu g\ ml^{-1}$ from *C. ternatea* L. and A. *lakoocha* Rox., respectively.

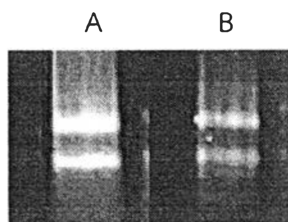
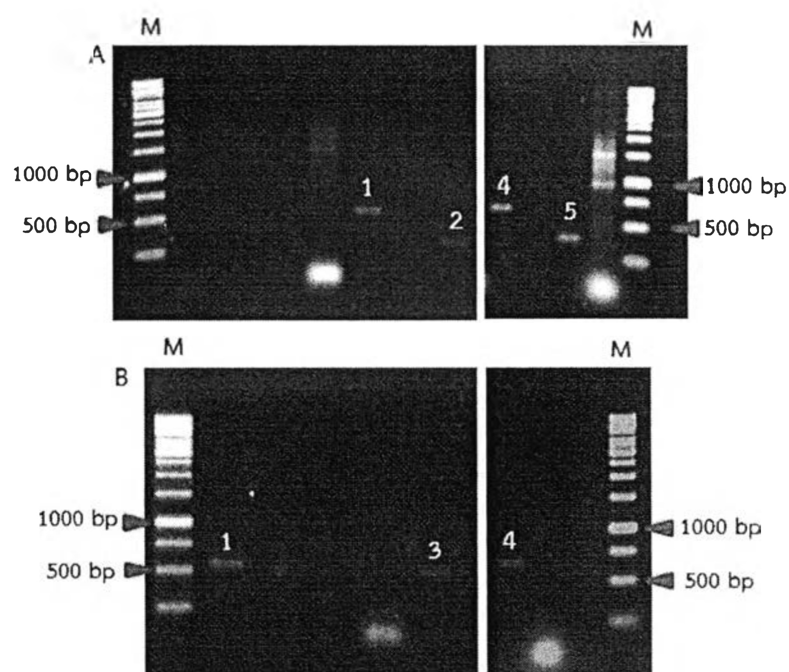


Figure 14 Agarose gel of the total RNA isolated from *C. ternatea* L. (A) and *A. lakoocha* Rox (B).

4.2 Isolation of core sequences from degenerate primers

Partial gene sequences were performed by PCR technique using the cDNA synthesized from RT-PCR as a template with multiple pairs of the degenerate primers. The results are showed in Figure 15. The pairs of degenerate primers F2R1, F2R3, F6R1 and F6R3 were able to amplify partial gene sequences from cDNA of *C.*

ternatea while the partial gene sequences from *A. lakoocha* were amplified by F2R1, F3R1 and F6R1.



2747256186

Figure 15 Agarose gel of the partial gene sequences from *C. ternatea* L. (A) and *A. lakoocha* Rox (B) amplified by multiple pairs of the degenerate primers. M: 1 kb DNA marker, Band number 1, 2, 3, 4, and 5 represented the partial gene sequences from sets of primers including F2R1 (544 bp), F2R3 (283 bp), F3R1 (479 bp), F6R1 (640 bp), and F6R3 (414 bp), respectively.

4.3 Full length genes from RACE PCR

Based on partial gene sequences of *C. ternatea* obtained from F6R1 primer pair, the new set of primers were designed overlapping the region between upstream and downstream of 5'-end (CTin_R, CTout_R, ALRin_R and ALRout_R) and 3'-end (RACEin_F and RACEout_F). The contig of 5'-RACE PCR was obtained from *C. ternatea* (1000 bp) and *A. lakoocha* (500 bp). In 3'-RACE PCR, the contig was obtained from *C.*

ternatea (700 bp) and *A. lakoocha* (900 bp) as showed in Figure 16. Then all contigs from both ends were aligned with the isolated partial gene sequence and the putative full-length cDNAs were recovered. The open reading frame (ORF) was determined from the start codon of ATG to the stop codon of TAA by ClonManager 9.0. To retrieve the full coding sequence cDNAs, the primers covering the whole coding sequence were designed from start and stop codons and amplified by PCR.

The full length cDNA of *ctl* gene, 1,495 bp, contained 122 bp at 5' untranslated region and 149 bp of 3' untranslated region with a poly (A) tail. The *ctl* showed open reading frame (ORF) of 1,224 bp (Figure 17A) encoding a putative polypeptide of 407 amino acids residues (Figure 18). For *alrc2* gene sequence retrieved by RACE, the 1,625 bp full length cDNA of *alrc2* gene was obtained and it contained 134 bp and 210 bp of 5' untranslated region and 3' untranslated region with a poly A tail, respectively. The *alrc2* showed open reading frame of 1,233 bp (Figure 17B) encoding a putative polypeptide of 410 amino acids residues (Figure 19).

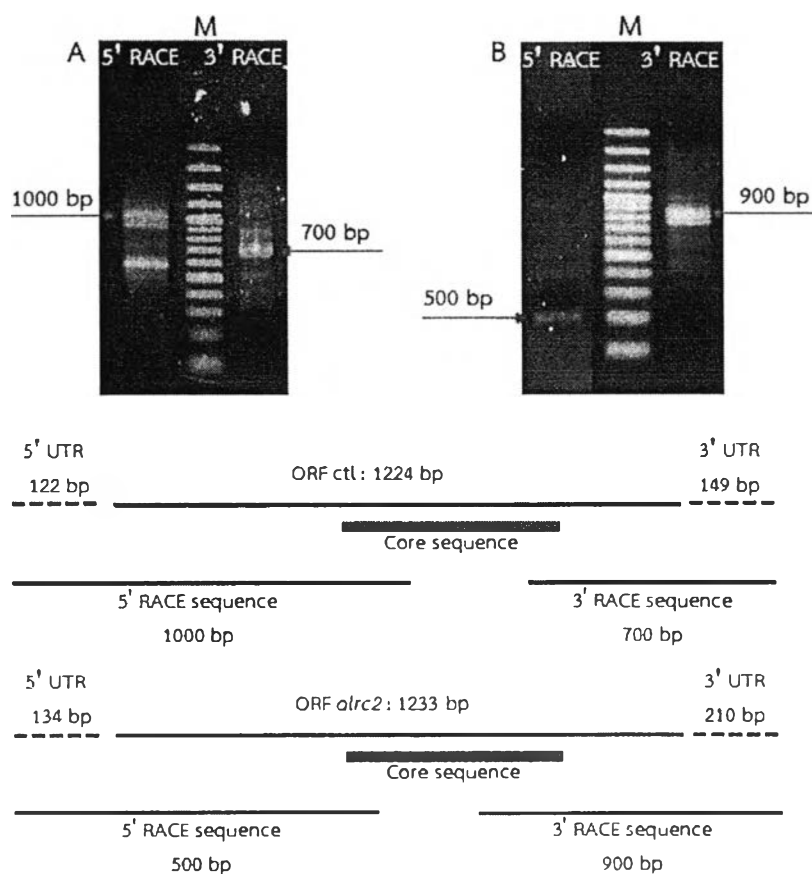


Figure 16 Isolation of full length cDNA of ctl and alrc2 by RACE PCR. The nested RACE-PCR products (5' and 3' fragments) from *C. ternatea* (A) and *A. lakoocha* (B) are shown on 1% agarose gel, M: 1 kb DNA marker. The contigs from RACE-PCR products and core sequences were aligned to obtain the full-length sequence of *ctl* and *alrc2* (C).

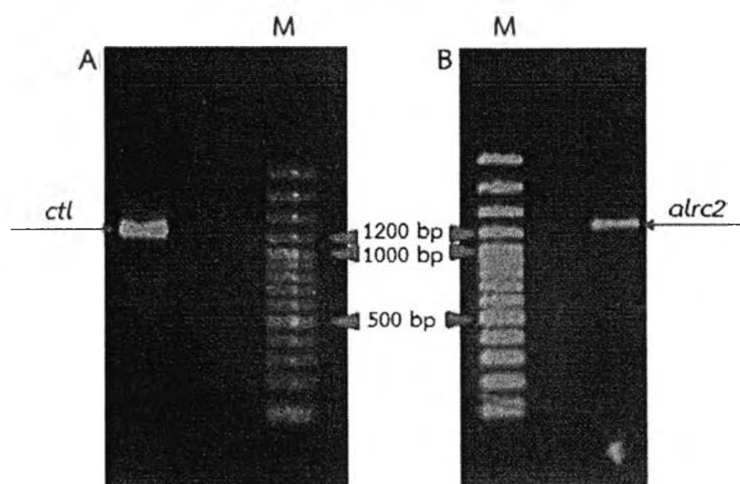


Figure 17 Agarose gels of *ctl* and *alrc2* coding sequence. The bands are shown on 1% agarose gel, M: 100 kb DNA marker and *ctl* and *alrc2* coding sequences amplified from *C. ternatea* (A) and *A. lakoocha* (B), respectively.

2747256186

```
   1  GACAGTGTGG TTCTCAAATA TCTGTGTTTC AGTGAATCGG CGGCGGAGGC GAGGGCGAAG GTTTAGGGAA AGAAAGAGTG
  81  ATTGTGGTTG TGAGCTTTTG AGTAGCAGTT TACAGTAGTC ACATGGATTC GGTGCTCTAT GGATCTTTGC CTAAGGCCTC
   1                                                M  D   S  V  L  Y   G  S  L   P  K  A
 161  TTCACTAACC ACTGGTGCCA ATTTCTGGAC TACTAAATGT CGTGCCCACA ATTACCATGC AAGCTCTTAT GCACCAAAAG
  13  S  S  L  T   T  G  A   N  F  W   T  T  K  C   R  A  H   N  Y  H   A  S  S  Y   A  P  K
 241  CCTCATGGCA CAAATGGAAA TTCCATAAAG AATACAGTGT TTTAAGGTTT AGACAATCAA GCTTGAGCCA TCATTACAAA
  40  A  S  W   H  K  W  K   F  H  K   E  Y  S   V  L  R  F   F  Q  S   S  L  S   H  H  Y  K
 321  GGCATTGGCG GAGGGTCTAC ACATCAAGAA AGTAACAGGA GATATGTTGT GAAAGCGGCC TCTGGACAAT CTTTTGAATC
  67  G  I  G   G  G  S   T  H  Q  E   S  H  K   A  Y  V   V  K  A  A   S  G  Q   S  E  L
 401  TGAACCCCAA GCTTTTGATC AGAAAAGCAT TTTGGACTCT GTCAAAAATT CCTTGGATGC TTTCTACAGG TTTTCTAGGC
  93  S  E  P  Q   A  F  D   Q  K  S   I  L  D  S   V  K  N   S  L  D   A  F  Y  R   F  S  R
 481  CACACACAGT TATTGGCACA GCATTAAGCA TAATTTCTGT ATCTCTCCTT GCATTGGAGA AATTATCTGA TATATCTCCA
 120  P  H  T   V  I  G  T   A  L  S   I  I  S   V  S  L  L   A  L  E   K  L  S   D  I  S  E
 561  ATGTTTTTTA CTGGTGTGTT GGAGGCTGTG GTTGCTGCCC TGTTTATGAA TATTTATATT GTTGGTTTGA ATCAATTATC
 147  M  F  F   T  G  V   L  E  A  V   V  A  A   L  F  M   N  L  Y  I   V  G  L   N  Q  L
 641  TGATGTTGAA ATAGACAAGA TAAACAAGCC ATATCTTCCA CTGGCATCCG GAGAATACTC CTTTGGAACT GGTGTTACTA
 173  S  D  V  E   I  D  K   I  N  K   P  Y  L  P   L  A  S   G  E  Y   S  F  G  T   G  V  T
 721  TTGTTGCATC ATTTTCAGTT CTGAGTTTTT GGCTTTGCTG GATTGTAGGT TCATGGCCAT TGTTTTGGGC TCTTTTTGTC
 200  I  V  A   S  F  S  V   L  S  F   W  I  C   W  I  V  G   S  W  E   L  F  W   A  L  F  V
 801  AGTTTTGTGC TAGGGACTGC TTATTCAATC AATGTGCCCC TTTTAAGATG GAAGAGGTTT GCAGTGCTTG CAGCAATGTG
 227  S  F  V   L  G  T   A  Y  S  I   N  V  P   L  L  R   W  K  R  F   A  V  L   A  A  M
 881  CATTCTAGCT GTTCGTGCAG TAATAGTTCA ACTTGCATTT TTCCTGCACA TGCAGACCCA TGTGTACAAG AGGCCAGCTG
 253  C  I  L  A   V  R  A   V  I  V   Q  L  A  F   F  L  H   M  Q  T   H  V  Y  K   R  P  A
 961  TCTTCTCAAG ACCTCTGATT TTTGCTACTG CATTCATGAG CTTCTTCTCT GTAGTTATAG CATTGTTCAA GGATATACCT
 280  V  F  S   R  P  L  T   F  A  T   A  F  M   S  F  F  S   V  V  I   A  L  F   K  D  I  E
1041  GACATTGAAG GGGATAAAAT ATTTGGCATC CAATCCTTTT CAGTACGTTT AGGTCAGAAG CGGGTATTCT GGATCTGTGT
 307  D  I  E   G  D  K   I  F  G  I   Q  S  F   S  V  F   L  G  Q  K   R  V  F   W  I  C
1121  TTCCCTTCTT GAAATAGCTT ATGGAGTCGC CCTCATGGTG GGAGCAGCAT CTCCCTGTCT CTGGAGTAAA GCTATCACGG
 333  V  S  L  L   E  I  A   Y  G  V   A  L  M  V   G  A  A   S  P  C   L  W  S  K   A  I  T
1201  GTGCGGGACA TGCTGTTCTG GCTTCACTTC TCTGGTATCA GGCCAAATCT GTAGATTTGA ATACCAAAGC TTCGATAACA
 360  G  A  G   H  A  V  L   A  S  L   L  W  Y   Q  A  K  S   V  D  L   N  T  K   A  S  I  T
1281  TCGTTCTACA TGTTTATCTG GAAGCTATTT TACGCAGAAT ACCTGCTCCT CCCTTATGTT AGATGAGGAT GCAGGGGCTT
 387  S  F  Y   M  E  I   W  K  L  F   Y  A  E   Y  L  L   L  F  Y  V   R  -
1361  TGTTGACTTT AGATATACTT GTGTTCCAAA GGATGCTGCC TGTCACAGGC CGGGCCTGGT GTCTGCACAA GTTTTAAGTT
1441  TTTCACAGCA ATTGTAAATG AAGAATTACT TTTGGGATTA AAAAAAAAAA AAAAA
```

Figure 18 The full length cDNA of *ctl* gene and its translated protein. The translated amino acids were decoded from ORF of *ctl* and indicated below their corresponding nucleotide codon.

2747256186

```
1     AAAACCCCAA AAACAGAAGT ATATTCAGAA ACCAGTCTCG CTCCTTTTGC AAGTACAAGA GTGGGCAAGA GTGTGAATTT

81    TTCCTTTTCG TTTTTTTTTT TTTTTTTCTT TTTGCTTGTT CTAATACATT GCCAATGGAT TCTTTTCTTC TGGGTTCATT
                                                                >>........ALRC2..........>
                                                                 M  D  S  F  L  L  G  S

161   GAAAGGTCCT TCTCTACTAG CAAATGGAGT GAATCATTGG AAGAAGGACA ATCTCAAGAA GGTTCGCTTT TCAGGTTCAT
      >....................................ALRC2......................................>
       L  K  G  F  S  L  L  A  N  G  V  N  H  K  K  K  D  N  L  K  K  V  R  F  S  G  S

241   ATGTATCACA TAGTCCTTCG AGCTTCAGTG AGTGGAACGT CATAGAAAGA CGATGCACTG CTAGGTTTCA ACATGCTCTT
      >....................................ALRC2......................................>
       Y  V  S  H  S  F  S  S  F  S  E  K  N  V  I  E  R  R  C  T  A  R  F  Q  H  A  L

321   CCAAAGCATT GCACGAAGGG TACTAGAGAA GCTTCTACAT TTTACCATGG ACAAAACAAA AGACTCCTGG TAAATGCCAC
      >....................................ALRC2......................................>
        F  K  H  C  T  K  G  T  R  E  A  S  T  F  T  H  G  Q  N  K  R  L  L  V  N  A

401   TGCTGGACAC CCGCTCGAGT CGGAACCTGG AGCCACTAGT TCAAAAAATG CTTGGAACTC TACTAAAGGT GCCCTAGGTG
      >....................................ALRC2......................................>
       I  A  G  H  F  L  E  S  E  F  G  A  T  S  S  K  N  A  K  N  S  T  K  G  A  L  G

481   TTTTTTACAG GTTTTCTCGG CCACACACAG TCATAGGCAC AGCATTGAGC ATAGTCTCAG TTTCCCTCCT TGCAGTACAG
      >....................................ALRC2......................................>
       V  F  Y  R  F  S  R  F  H  T  V  I  G  T  A  L  S  I  V  S  V  S  L  L  A  V  Q

561   AGACTTTCAG ATATTTCTCC ATTATTTTTC ACTGGGGTGC TGGAGGCTGT GGTTGCTGCC CTCTTTATGA ATATATATAT
      >....................................ALRC2......................................>
        R  L  S  D  I  S  F  L  F  F  T  G  V  L  E  A  V  V  A  A  L  F  M  N  I  Y

641   TGTTGGTTTG AATCAATTGT ATGACATTGA TATAGACAAG GTTAACAAGC CATCTCTTCC ATTGGCTTCA GGGGAATATT
      >....................................ALRC2......................................>
       I  V  G  L  N  Q  L  Y  D  I  D  I  D  K  V  N  K  F  S  L  F  L  A  S  G  E  Y

721   CCGTTTCAAC TGGCACCTTG ATTGTCACAT CCTTCGCTGT TCTGAGCTTT TGCCTCTCAT GGATCGTTGG TTCATGGCCC
      >....................................ALRC2......................................>
       S  V  S  I  G  T  L  I  V  T  S  F  A  V  L  S  F  C  L  S  W  I  V  G  S  K  F

801   TTGTTTTGGG CGCTCTTCAT AAGTTTCGTA CTTGGAACTG CTTATTCAAT CAATATGCCC CTTTTGAGAT GGAAGAGATT
      >....................................ALRC2......................................>
        L  F  W  A  L  F  I  S  F  V  L  G  T  A  T  S  I  N  M  P  L  L  R  W  K  R

881   TGCTGTAGTT GCTGCGATGT GCATCCTTGC GGTCCGTGCA GTGATTGTTC AACTAGCATT TTTCCTGCAC ATGCAGACCC
      >....................................ALRC2......................................>
       F  A  V  V  A  A  M  C  I  L  A  V  R  A  V  I  V  Q  L  A  F  F  L  H  M  Q  T

961   ATGTGTACAA AAGACCTGCC ATCTTCTCAA GGCCTCTGAT TTATGCTACT GCATTTATGA GCTTCTTCTC AGTTGTTATT
      >....................................ALRC2......................................>
       H  V  Y  K  R  F  A  I  F  S  R  F  L  I  Y  A  T  A  F  M  S  F  F  S  V  V  I

1041  GCATTGTTTA AGGATATACC TGATATCGAC GGAGATAGGA TATATGGTAT TCGATCTTTT ACAGTGCGGT TAGGTCAAAA
      >....................................ALRC2......................................>
        A  L  F  K  D  I  F  D  I  D  G  D  R  I  Y  G  I  R  S  F  T  V  R  L  G  Q

1121  GAAGGTATTC TGGATCTGCA TTTCACTTCT TGAAATGGCT TATAGCGTTG CTCTTTTAGT GGGGGCGTCA TCTGGTTTCT
      >....................................ALRC2......................................>
       K  K  V  F  W  I  C  I  S  L  L  E  M  A  Y  S  V  A  L  L  V  G  A  S  S  G  F

1201  TATGGAGCAA AGTTGCTACG GTTTTGGGAC ACACAATCTT GGCTTCATTA CTTTGGGGTC GTGCCAAGTC CGTTGATTTG
      >....................................ALRC2......................................>
       L  W  S  K  V  A  T  V  L  G  H  T  I  L  A  S  L  L  W  G  R  A  K  S  V  D  L

1281  TCCAGCAAAG CTGCAATAAC ATCCTTTTAC ATGTTTATAT GGAAGCTCTT TTATGCCGAG TATCTACTTA TACCGCTCGT
      >....................................ALRC2......................................>
        S  S  K  A  A  I  T  S  F  Y  M  F  I  W  K  L  F  Y  A  E  Y  L  L  I  P  L

1361  TAGGTGAAAG AAATAGAGAC AGAATGTTGT ATAAAGGGTA TTTATAGGTT TGAGTTTATT TGACGGATAA TCTAAGAAGG
      >....>> ALRC2
       V  K  *

1441  GATAGAAAAA GTTTATTTGG GAAGTGATTT AGTGTTGGAT GTTCAAAGTA GGGATTTAAG TTCATTTGGG AAGTGCAAAA

1521  TGTATCGTAT TGCCAAAGTC CTGAAAATGT GCTAATGTTG TACGTAAGAA GCAGCCTTGC TAAATGAGGA GTTGAGAAAC

1601  TAATTTGGGA AAAAAAAAAA AAAAA
```

Figure 19 The full length cDNA of *alrc2* gene and its translated protein. The translated amino acids were decoded from ORF of *alrc2* and indicated below their corresponding nucleotide codon.

4.4 Cloning of full length genes

The full length genes obtained from RACE-PCR were cloned into pGem-T easy vector and subjected to sequence verification. To confirm the insertion to the vector, the pGem-T vectors carrying the genes were cut by *Eco*RI restriction enzyme. After enzyme digestion, the band of full length *ctl* and *alrc2* genes along with the bands of pGemT easy backbone at 3 kb were clearly seen on the agarose gel (Figure 20). The full length gene sequences were confirmed by sequencing.

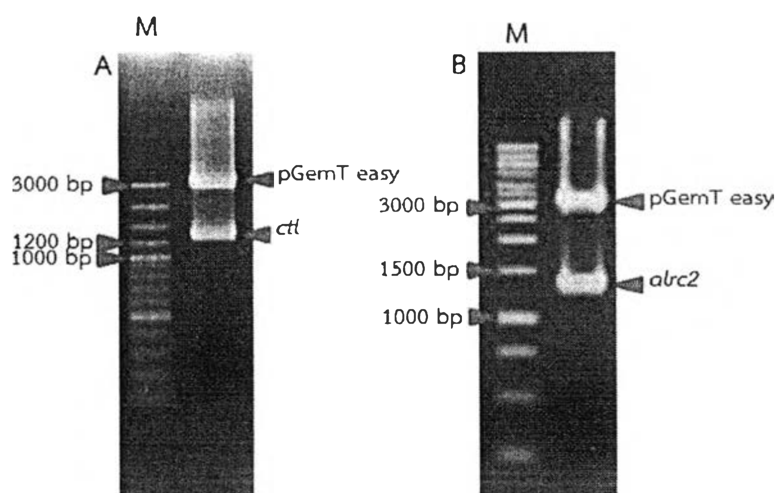


Figure 20 Verification of gene insertion to pGemT vector by restriction enzyme digestion are shown on 1% agarose gel against 1 kb DNA marker (M). The *EcoRI* restriction enzyme was used to digest recombinant pGem-T easy vector of *ctl* (A) and *alrc2* (B).

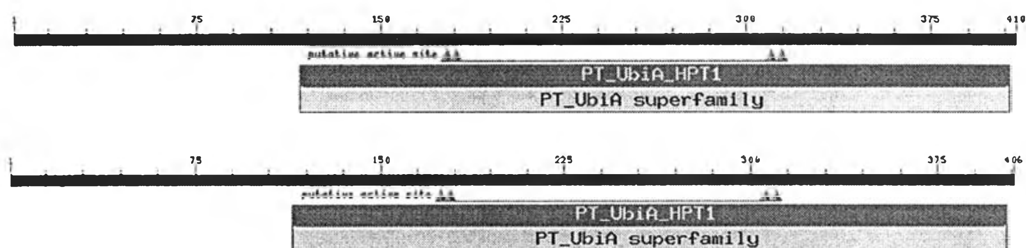
4.5 *In silico* protein identification and characterization

All deduced proteins from ORF of *ctl* and *alrc2* were predicted for their functional domain group by PSI-blast search (http://www.ebi.ac.uk/Tools/sss/psiblast/) against the UniProtKB/Swiss-Prot database.

2747256186

The deduced proteins showed their significant E-value with UbiA prenyltransferase family of homogentisate prenyltransferase (Figure 21). Then the sequences were predicted for their molecular weights and theoretical pI values by the Compute pI/Mw Tool from ExPASy (http://web.expasy.org/compute_pi/). The predicted sizes were 45.58 and 45.59 kDa and pI values were 9.53 and 9.87 of CTL and ALRC2, respectively (Table 12). The results indicated that *ctl* and *alrc2* are in the group of UbiA prenyltransferase gene family and possibly encoded putative proteins functioning as homogentisate prenyltransferase.

2747256186



Name	Accession	Description	Interval CTL	Interval ALRC2	E-value CTL	E-value ALRC2
PT_UbiA _HPT1	cd13960	Tocopherol phytyltransferase	114-403	117-407	9.64E-153	2.08E-148
PLN0287 8	PLN02878	homogentisate phytyltransferase	128-406	131-410	0.00E+00	0.00E+00
UbiA	PRK12887	tocopherol phytyltransferase	111-403	102-407	2.89E-96	2.97E-99

Figure 21 PSI-blast search of the putative proteins CTL and ALRC2.

Table 12 Summary of computed pI and MW of the deduced proteins.

	ORF (bp)	Amino acid (residues)	pI	MW (kDa)
alrc2	1233	410	9.87	45.59
ctl	1224	407	9.53	45.58

The transmembrane (TM) domains of CTL and ALRC2 were predicted by TMHMM program (http://www.cbs.dtu.dk). These genes contained nine putative TM domains (Table 13 and Figure 22) and possessed a conserved prenyltransferase motif (NQXXDXXXD) and an aspartate rich motif (KD(I/L)XDX(E/D)GD) between TM domain 2 and 3 and TM domain 6 and 7, respectively (Figure 23). Subcellular localization of all amino acid sequences were predicted on their N-terminal peptide by TargetP 1.1. The result showed that this program was failed to assign the organelle localization with low reliable prediction (RC=5) as showed in Table 14. Therefore, the proteins were predicted localization and signal peptide position by other programs e.g. SignalP 4.1 (http://www.cbs.dtu.dk/services/SignalP/), WoLF PSORT (http://www.genscript.com/psort/wolf_psort.html) and Protcomp 9.0 (http://www.softberry.com/berry.phtml?topic=protcomppl). Using the SignalP, it was found that the amino acid sequences showed low signal peptide (S- score $\leq$ 0.2) as shown in Figure 24 while CTL and ALR were chloroplast protein containing transit peptide 15 and 17 amino acid residues when prediction by WoLF PSORT and Protcomp (Table 15).

2747256186

Table 13 List of transmembrane domains of ALRC2 and CTL.

	ALRC2	CTL
	Transmembrane alpha helix regions	
TM1	VIGTALSIVSVSLLAVQRL	VIGTALSIISVSLLALEKL
TM2	FFTGVLEAVVAALFMNIYIVG	FFTGVLEAVVAALFMNIYIVG
TM3	GEYSVSTGTLIVTSFAVLSFCLS	GEYSFGTGVTIVASFSVLSFWLC
TM4	SWPLFWALFISFVLGTAYSINMP	SWPLFWALFVSFVLGTAYSINV
TM5	VAAMCILAVRAVIVQLAFFLHMQ	LAAMCILAVRAVIVQLAFFLHMQ
TM6	RPAIFSRPLIFATAFMSFFSVVIALF	RPLIFATAFMSFFSVVIALF
TM7	ICISLLEMAYSVALLVGASS	ICVSLLEIAYGVALMVGAAS
TM8	KVATVLGHTILASLLWGRAKSV	KAITGAGHAVLASLLWYQAKSV
TM9	SFYMFIWKLFYAEYLLIPFVR	SFYMFIWKLFYAEYLLLPYVR

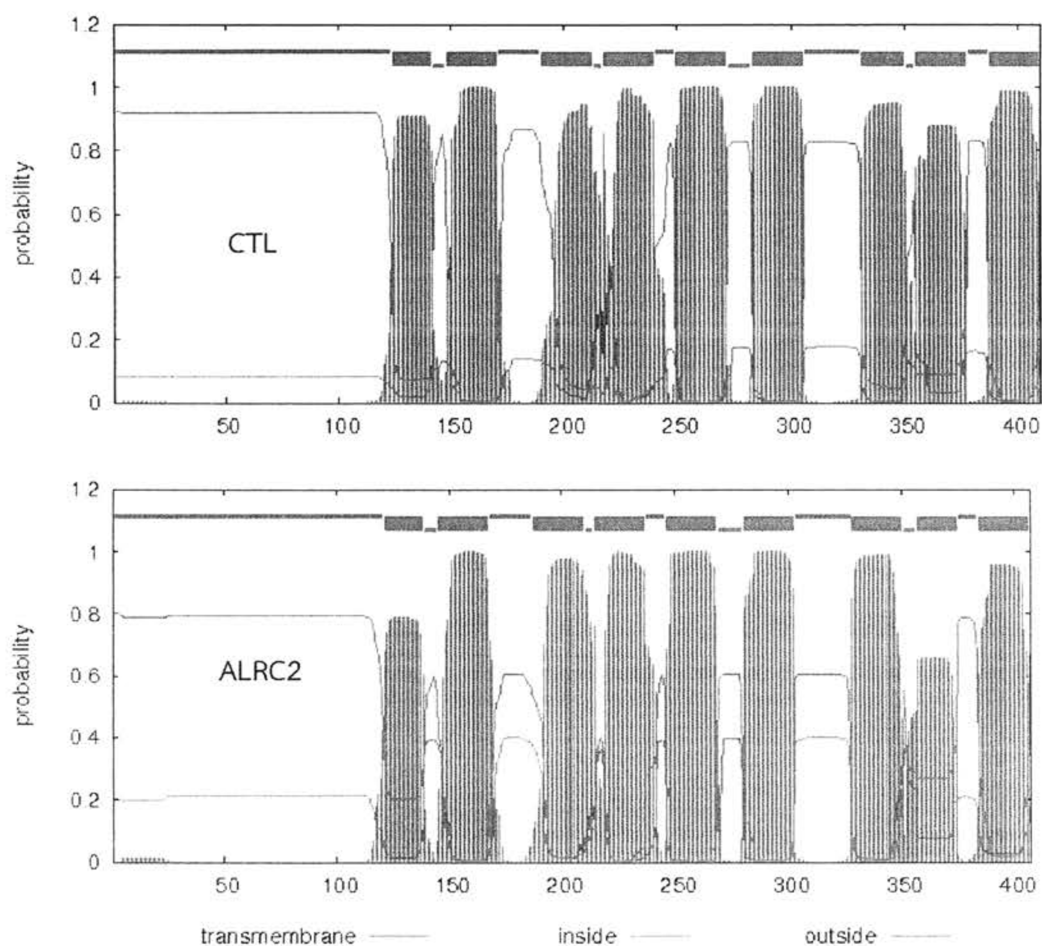


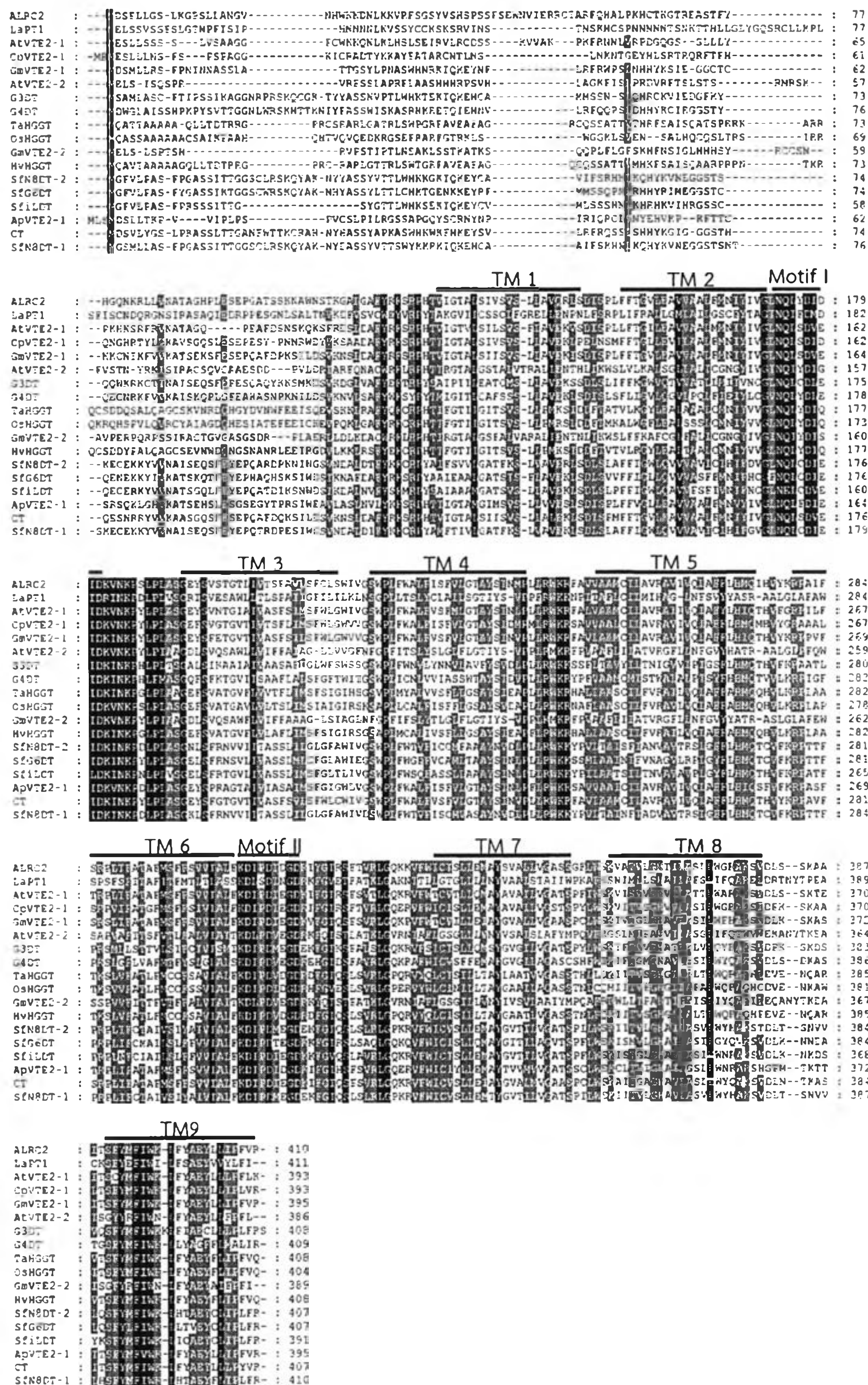
Figure 22 TMHMM analysis of ALRC2 and CTL protein sequences. The colors showed the probabilities for transmembrane regions (red), inside region of an organelle (blue) and outside of an organelle (pink).

Table 14 Sequence data analysis by TargetP.

Name	Len	cTP	mTP	SP	other	Loc	RC
ALRC2	410	0.422	0.073	0.030	0.460	*	5
CTL	407	0.074	0.444	0.023	0.626	*	5

Table 15 Sequence data analysis by WoLF PSORT and Protcomp.

	Amino acid sequence	Similarity in location DB	Extracellula r score	Intigral score	chloroplast transit peptide
CTL	407	chloroplast	0.9	9.3	15
ALRC2	410	chloroplast	0.9	9.5	17



2747256186

Figure 23 Multiple alignment of prenyltransferases family in plants. Motif I (NQxxDxxxD) and Motif II (KD(I/L)xDx(E/D)GD) indicated with black arrows are

conserved amino acid sequences among prenyltransferases and transmembrane domain (TM) are indicated with black lines.

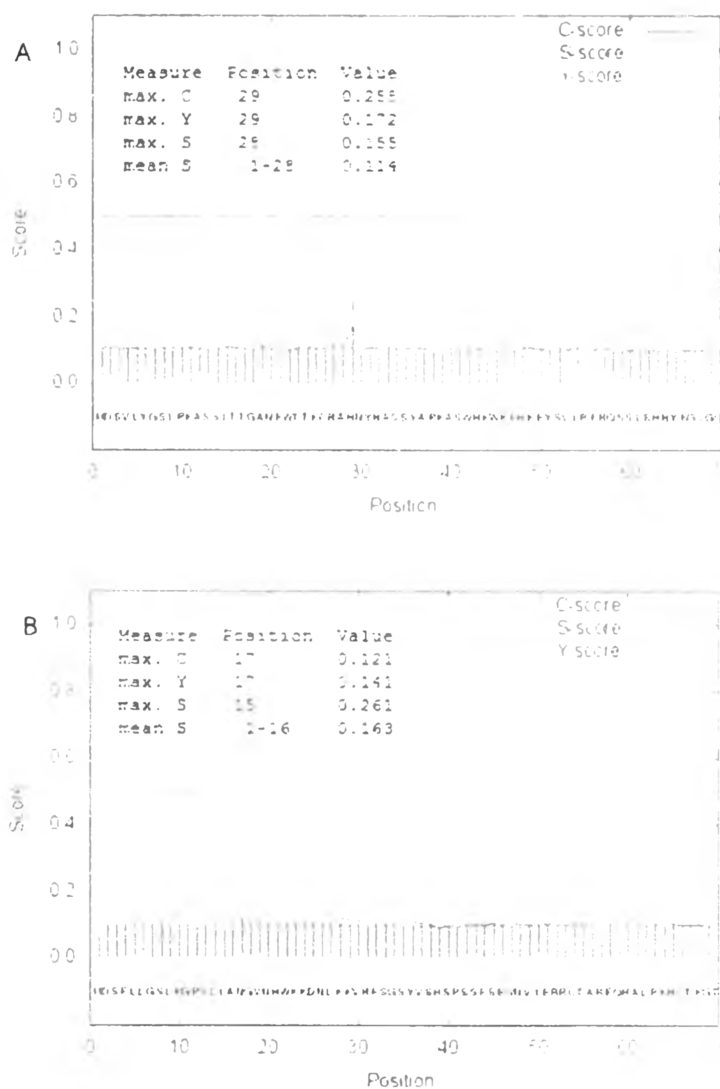


Figure 24 The graphical image of transmembrane prediction by SignalP, C-, S-, and Y-score cleavage site were predicted to be at position of maximal Y score (A) CTL and (B) ALRC2. C-score: raw signal peptide cleavage sites, S-score: positions within signal peptides from positions in the mature part of the proteins and Y-score: combined cleavage site score (C and S).

4.6 Phylogenetic analysis

Amino acid sequences of CTL and ALRC2 from candidate genes were compared to their homologs by using BLASTX algorithm in the NCBI database. The CTL and ALRC2 showed high sequence homology to homogentisate phytyltransferase of *Glycine max* and *Morus notabilis* with 84% and 88%, respectively. The phylogenetic tree of prenytransferase families was constructed including homogentisate phytyltransferase (HPT), flavonoid prenytransferase (flavonoid PTase), homogentisate geranylgeranyltransferase (HGGT), and homogentisate solanesyl transferase (HST) and shown in Figure 25. The analysis revealed that the proteins CTL and ALRC2 are closely related to the group of HPTs. The CTL and ALRC2 were highly similar to VTE2-1 involved in vitamin E biosynthesis in plants. Further sequence examination, the conserved aspartate rich regions (motif I and motif II) of various prenyltransferases were compared (Figure 26). In motif I, the CTL showed high similarity with the HPT of *G. max* (GmVTE2-1) and *Arabidopsis thaliana* (AtVTE2-1) while the ALRC2 were closest to the HPT of *Cuphea avigera* var. *pulcherrima* (CpVTE2-1) and *M. notabilis* (MnVTE2-1). In motif II, the CTL showed similarity with a group of HPTs of AtVTE2-1, GmVTE2-1, CpVTE2-1, *Triticum aestivum* (TaVTE2-1), and flavonoid prenyltransferase of *Sophora flavescens* (SfiLDT). The ALRC2 showed similarity with a group of HPTs of *Allium ampeloprasum* (ApVTE2-1), MdVTE2-1 and homogentisate geranylgeranyltransferase of *Oryza sativa Japonica* (OsHGGT). These results suggested that CTL and ALRC2 were likely to be HPT/VTE2 enzymes.

2747256186

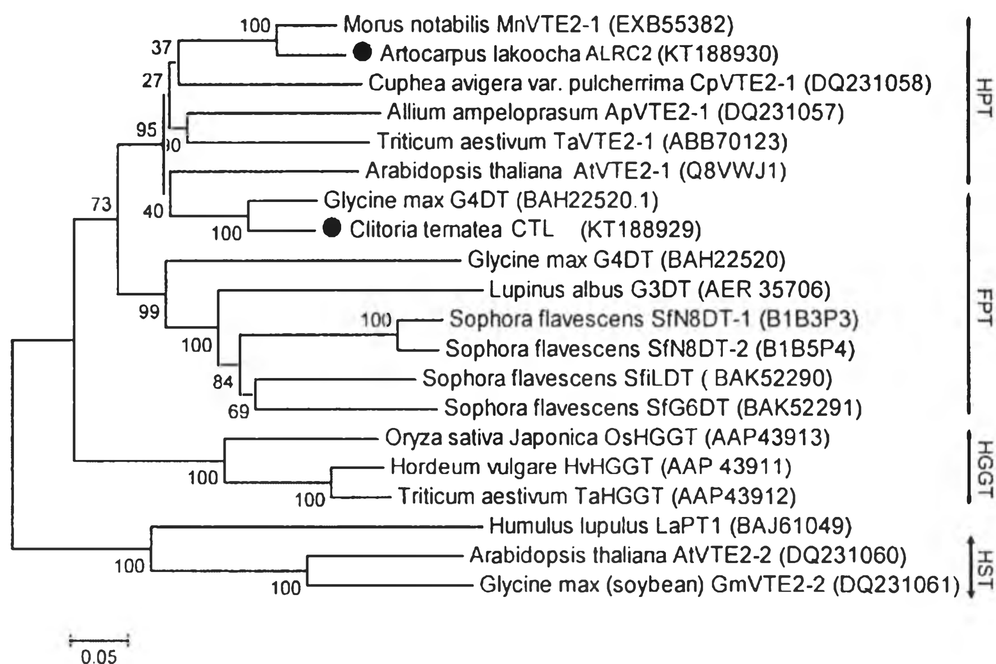
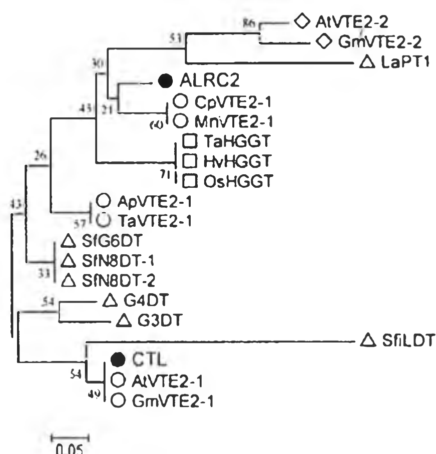


Figure 25 The phylogenetic tree of putative protein sequences of CTL and ALRC2 and related prenyltransferase proteins in plants. Protein sequences from various plant species were retrieved from NCBI database and their accession numbers were shown in parenthesis. The neighbor-joining was drawn using MEGA4. The optimal tree with the sum of branch length = 5.75502538 was shown. Bootstrap values 1000 replicate are shown, and the branch lengths represented relative genetic distances. The evolutionary distances were measured by JTT matrix-base method. The abbreviations were following: FPT, Flavonoid prenytransferase; HPT, homogentisate phytyltransferase; prenyltransferase and HGGT, homogentisate geranylgeranyltransferase.

2747256186

Motif I

AtVTE2-2 : NQIYDIGID
GmVTE2-2 : NQIYDISID
LaPT1 : NQIFDMDID
ALRC2 : NQLYDIDID
CpVTE2-1 : NQLSDIDID
MnVTE2-1 : NQLSDIDID
TaHGGT : NQLYDIQID
HvHGGT : NQLYDIQID
OsHGGT : NQLYDIQID
ApVTE2-1 : NQLFDIEID
TaVTE2-1 : NQLFDIEID
SfG6DT : NQLCDIEID
SfN8DT-1 : NQLCDIEID
SfN8DT-2 : NQLCDIEID
G4DT : NQLYDLEID
G3DT : NQLCDLEID
SfiLDT : NELCDVELD
CTL : NQLSDVEID
AtVTE2-1 : NQLSDVEID
GmVTE2-1 : NQLSDVEID



Motif II

GmVTE2-2 : KDLPDVEGD
AtVTE2-2 : KDLPDVEGD
ApVTE2-1 : KDIPDIDGD
MnVTE2-1 : KDIPDIDGD
OsHGGT : KDIPDIDGD
ALRC2 : KDIPDIDGD
HvHGGT : KDIPDVDGD
TaHGGT : KDIPDVDGD
SfG6DT : KDIPDTEGD
G3DT : KDIPDMEGD
SfN8DT-1 : KDIPDMEGD
SfN8DT-2 : KDIPDMEGD
CpVTE2-1 : KDIPDIEGD
SfiLDT : KDIPDIEGD
CTL : KDIPDIEGD
TaVTE2-1 : KDIPDIEGD
AtVTE2-1 : KDIPDIEGD
GmVTE2-1 : KDIPDIEGD
G4DT : KDIPDVEGD
LaPT1 : KDLSDINGD

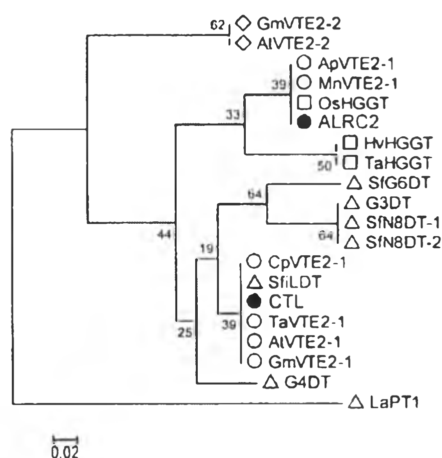


Figure 26 Phylogenetic trees for conserved amino acid sequences (the aspartate rich regions) of prenyltransferase family. The alignment of motif I (A) and motif II (B) from prenyltransferase family. Trees were generated by MEGA6 with neighbor-joining method. The available protein sequences from VTE, GDT, HGGT, and PT groups were retrieved from various species including *Arabidopsis thaliana* (At), *Allium ampeloprasum* (Ap), *Cuphea avigera* var. *pulcherrima* (Cp), *Glycine max* (Gm), *Hordeum vulgare* (Hv), *Humulus lupulus* (La), *Morus notabilis* (Mn), *Oryza sativa Japonica* (Os), *Triticum aestivum* (Ta) and *Sophora flavescens* (Sf).

2747256186

4.7 Construction of plant expression vectors

The ORF genes were amplified with sets of primers in 3.15.1. The PCR products were overhanging at start codon and blunt end at stop codon. Then the gel-purified PCR products were cloned into pENTR™/D-TOPO® vector by Topoisomerase I to produce the entry clone and transformed into *E. coli*. The obtained recombinants were checked for their insertional direction by *Not*I restriction digestion, PCR, and sequencing. The correct orientation pattern of recombinants showed 3.8 kb but empty vector showed 2.5 kb when cut with *Not*I restriction enzyme (Figure 27A and 27B). Amplification of *ctl* and *alrc2* from the recombinant vectors by PCR showed the expected bands at about 1.2 kb (Figure 27C and 27D). The gene sequences were confirmed again by sequencing. Hence, the correct entry vectors were subcloned separately into a destination clone (pGWB6) by Gateway® LR clonase® II to produce the plant expression clones. The pGWB6::*ctl*, and pGWB6::*alrc2* were finally obtained and confirmed the recombinant vectors by PCR showed the expected bands at about 1.2 kb (Figure 28A and 28B), respectively. Each construct was subsequently transformed into *E. coli*. and *Agrobacterium tumefaciens* for further used in plant transformation.

2747256186

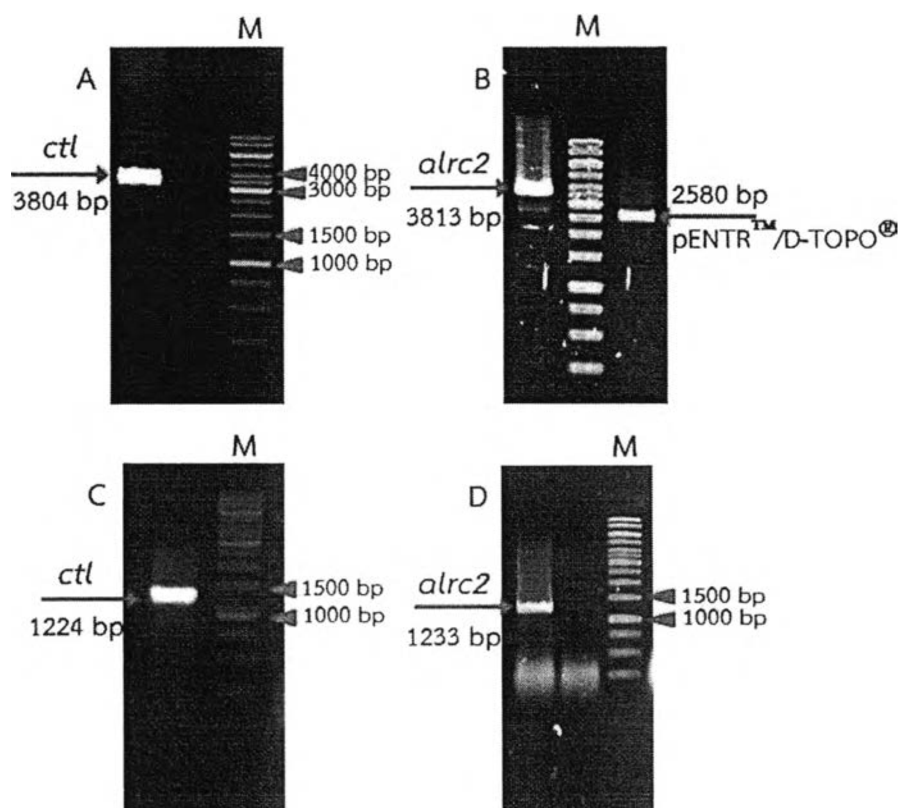


Figure 27 The construction of the entry vector (pENTR™/D-TOPO® harboring the *ctl* and *alrc2* genes were analyzed on 1% agarose gel against 1 kb DNA marker (M). The recombinant pENTR™/D-TOPO®:*ctl* digested with *Not*I restriction enzyme (A) and the recombinant pENTR™/D-TOPO®:*alrc2* digested with *Not*I restriction enzyme and the empty pENTR™/D-TOPO® vector (B). PCR products from recombinant *ctl* and *alrc2* in pENTR™/D-TOPO® vector (C and D).

2747256186

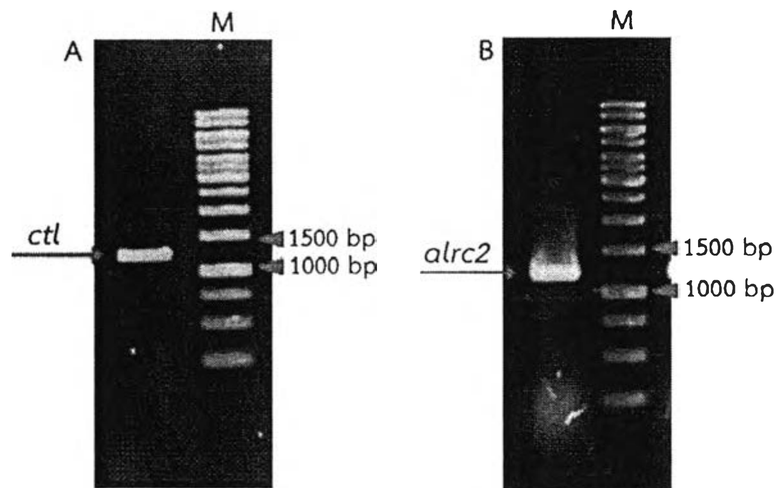


Figure 28 The construction of the destination vector. The PCR products amplified from recombinant *ctl* (A) and *alrc2* (B).in pGWB6 vector and shown on 1% agarose gel against 1 kb DNA marker (M).

2747256186

4.8 Gene expression of *ctl* and *alrc2* overexpressed in tomato leaves

The *ctl* and *alrc2* genes were transiently expressed in tomato leaves via *Agrobacterium*-mediated transformation. After infiltration of the recombinant expression vectors, leaves were harvested at 1, 3, 6 and 9 day post agroinfiltration; dpa and their total RNA were extracted for cDNA synthesis. During 1 – 6 dpa the visible phenotypes of infiltrated tomato leaf of empty vector and pGWB6::*ctl* not different when compare with control but these transient expressions were slightly induce cell death at 9 dpa that found light yellow around of brown spot. In transient of pGWB6:*alrc2* leaf showed visible phenotype same as control during 1 – 3 dpa but the leaf was induced cell death at 6 dpa that seem little yellow and increasing at 9 dpa. (Figure 29)

Expressions of *ctl* and *alrc2* were under the control of CaMV 35 promoter and the gene expression profile was performed by RT-PCR was performed on tomato

leaves overexpressed the genes as showed in Figure 30. The recombinant pGWB6::*ctl* showed highest expression at 1 dpa and sharply decreasing after 3 dpa according to RT-PCR (Figure 30A). The RT-PCR results showed that the recombinant pGWB6:*alrc2* showed expression between 1 and 6 dpa and very low at 9 dpa (Figure 30B).

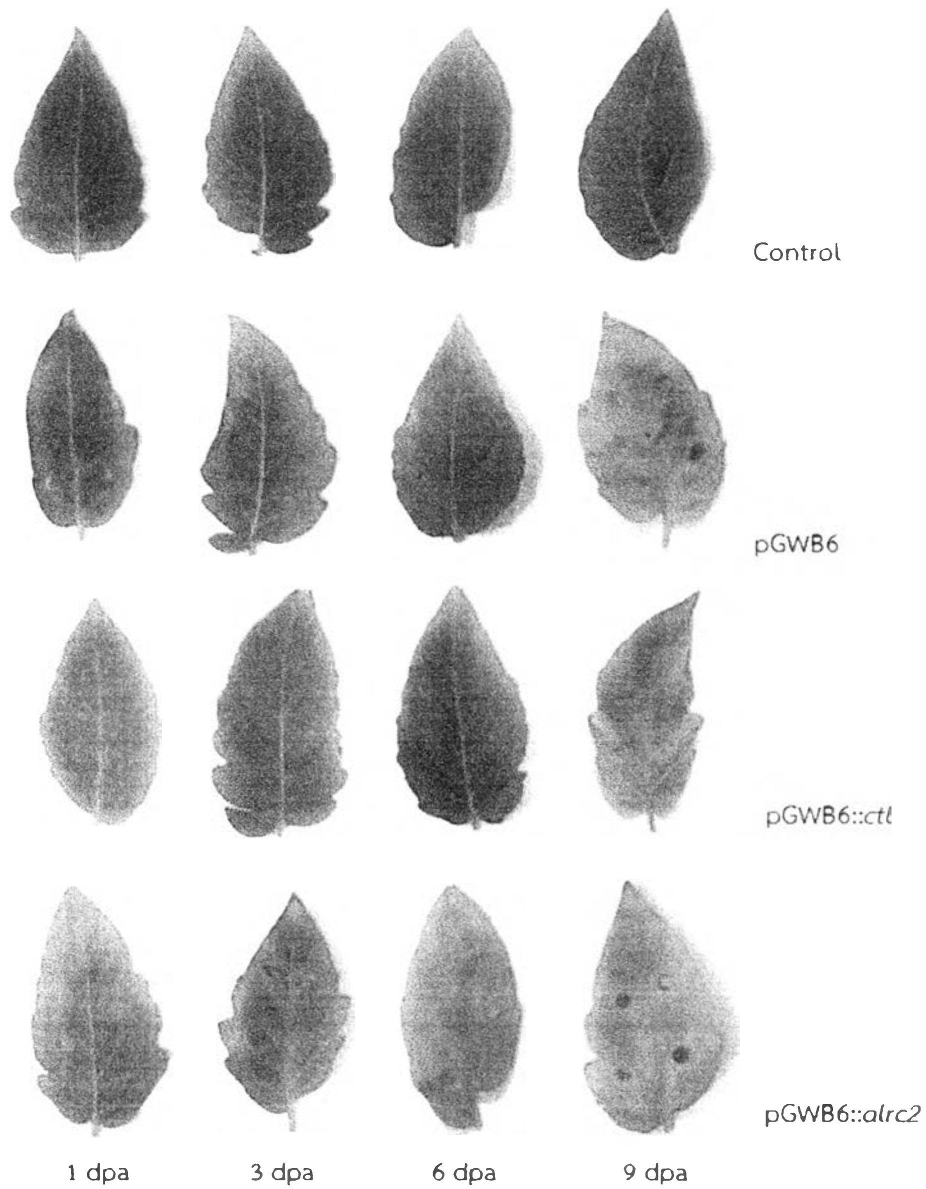


Figure 29 Tomato leaves after infiltration of the recombinant expression vectors via A. tumefaciens-mediated transformation. Leaves were harvested at 1, 3, 6 and and 9 day post agroinfiltration (dpa). Leaves were harvested at 1, 3, 6 and and 9 day post agroinfiltration (dpa).

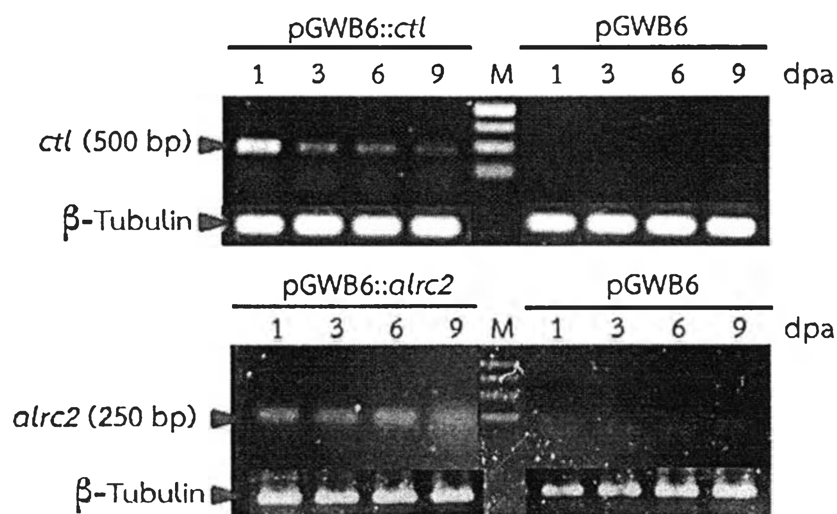


Figure 30 RT-PCR expression analysis of *ctl* and *alrc2* in the agroinfitrated tomato leaves at 1 – 9 dpa. β-Tubulin was served as an internal reference gene and the empty vectors were served as negative control. Gene Ruler 1 kb DNA ladder (M) was used to indicate the product size. The gene expression of the recombinant pGWB6::*ctl* (A) and pGWB6::*alrc2* (B) were detected on the 1 % agarose gel.

2747256186

4.9 Recombinant protein expression in tomato leaves

All proteins (CTL and ALRC2) in the pGWB6 vector expressed in transformed tomato leaves were fused with GFP protein (26.8 kDa) at the N-terminal of the proteins. Total proteins were extracted from the leaves and determined the protein concentration by Bradford's method. One hundred micrograms of total protein were loaded and separated on 10% SDS-PAGE gel. The proteins were detected by blotting with anti-GFP antibody conjugated with HRP on PVDF membrane and visualized using chemiluminescent HRP substrate (Figure 31). The CTL was slightly expressed at 1 dpa and gradually increased at 3 dpa. In contrary, the ALRC2 could not be detected at 1 dpa. However, it was gradually expressed at 3 and 6 dpa.

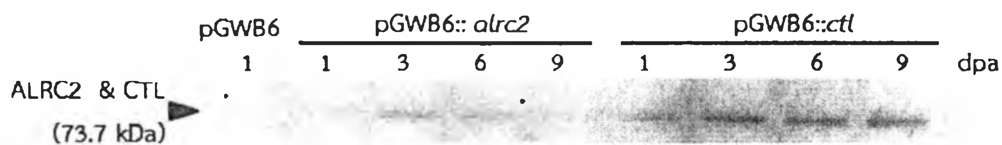


Figure 31 Detection of the recombinant proteins by western blots analysis. Expression of ALRC2 and CTL proteins from agroinfiltration tomato leaves using antibody against GFP protein in different dpa compared with the empty vector (pGWB6).

4.10 Determination of tocopherol content in agroinfiltrated tomato leaves

Transformed tomato leaves at 1, 3, 6 and 9 days were extracted and analyzed for the accumulation of α-tocopherol against control by TLC technique. The amount of α-tocopherol was determined against the α-tocopherol standard curve as shown in Figure 32. The recombinant protein from pGWB6::*ctl* showed high effectiveness in enhancing the production of α-tocopherol content at 3 dpa. The results showed 2.4 ± 0.38 fold increment of α-tocopherol content compared with the control (Figure 33) and the intensity of α-tocopherol band on TLC clearly showed the difference after 3 dpa (Figure 34A). Moreover, the transformation of pGWB6::*alrc2* could induce the accumulation of α-tocopherol in infiltrated tomato leaves at 3 dpa (Figure 34B). The α-tocopherol content was increased 1.4 ± 0.05 fold higher than the control (empty vector; pGWB6) (Figure 33).

4.11 Determination of total chlorophyll content in agroinfiltrated tomato leaves

When *ctl* and *alrc2* were introduced to tomato leaves mediate agrobacterium, the resulting transient plants showed the decreased total chlorophyll when compare with the control. The total chlorophyll of transient plants *ctl* and

2747256186

alrc2 dropped from 31.2 ± 0.34 to 28.2 ± 0.09 μg ml^{-1} and 32.1 ± 0.07 to 30.8 ± 0.06 μg ml^{-1}, respectively in order of day after agroinfiltration (1 – 9 dpa) that correlated with α-tocopherol accumulation (Figure 33).

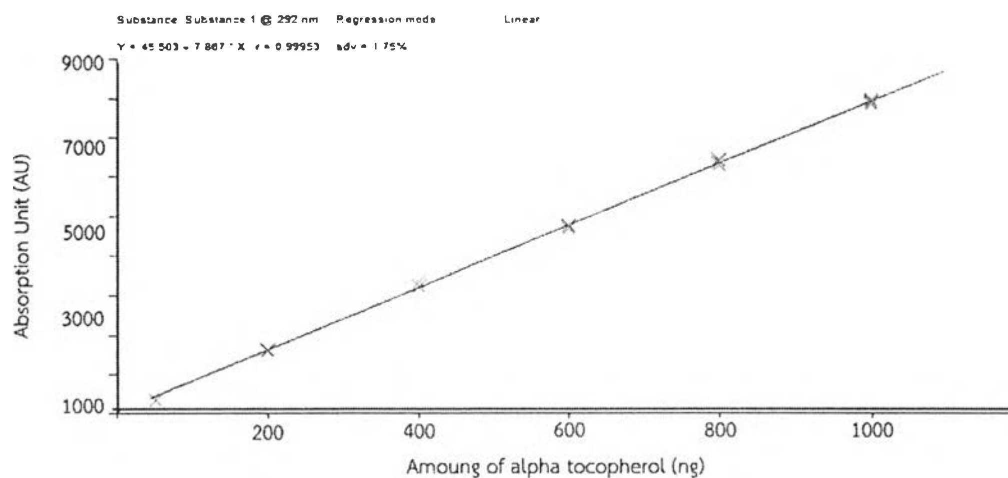


Figure 32 The standard curve of α-tocopherol. The amount of a-tocopherol was plotted against absorption unit (AU) measured on TLC plate developed with chloroform:cyclohexane (11:9 v/v) and scanned under 292 nm (n=3).

2747256186

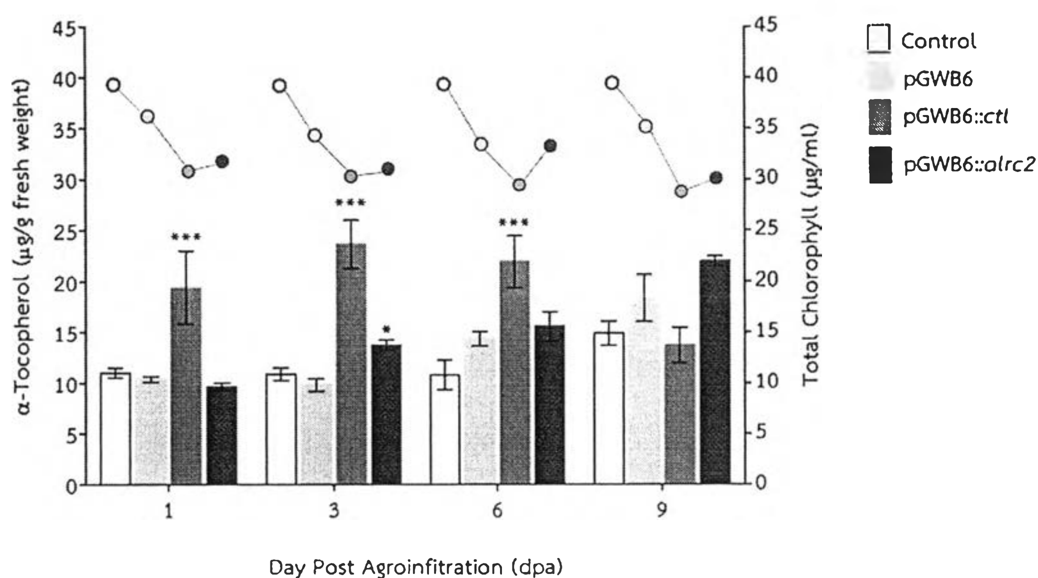


Figure 33 The α-tocopherol and total chlorophyll contents in pGWB6::*alrc2* and pGWB6::*ctl* agroinfiltrated leaves. Determination of α-tocopherol content (bar graph) was from the band intensity on TLC plate. The α-tocopherol content from agroinfiltrated leaves overexpressing *ctl* and *alrc2* genes were compared with empty vector (pGWB6) in 1 – 9 dpa. Data represent the mean and SD (n=3). The different between samples measurement by two way ANOVA test (*: $p < 0.05$, **: $p < 0.01$ and ***: $p < 0.001$). The total chlorophyll content (line graph) was measured by absorption at 663 and 645 nm.

2747256186

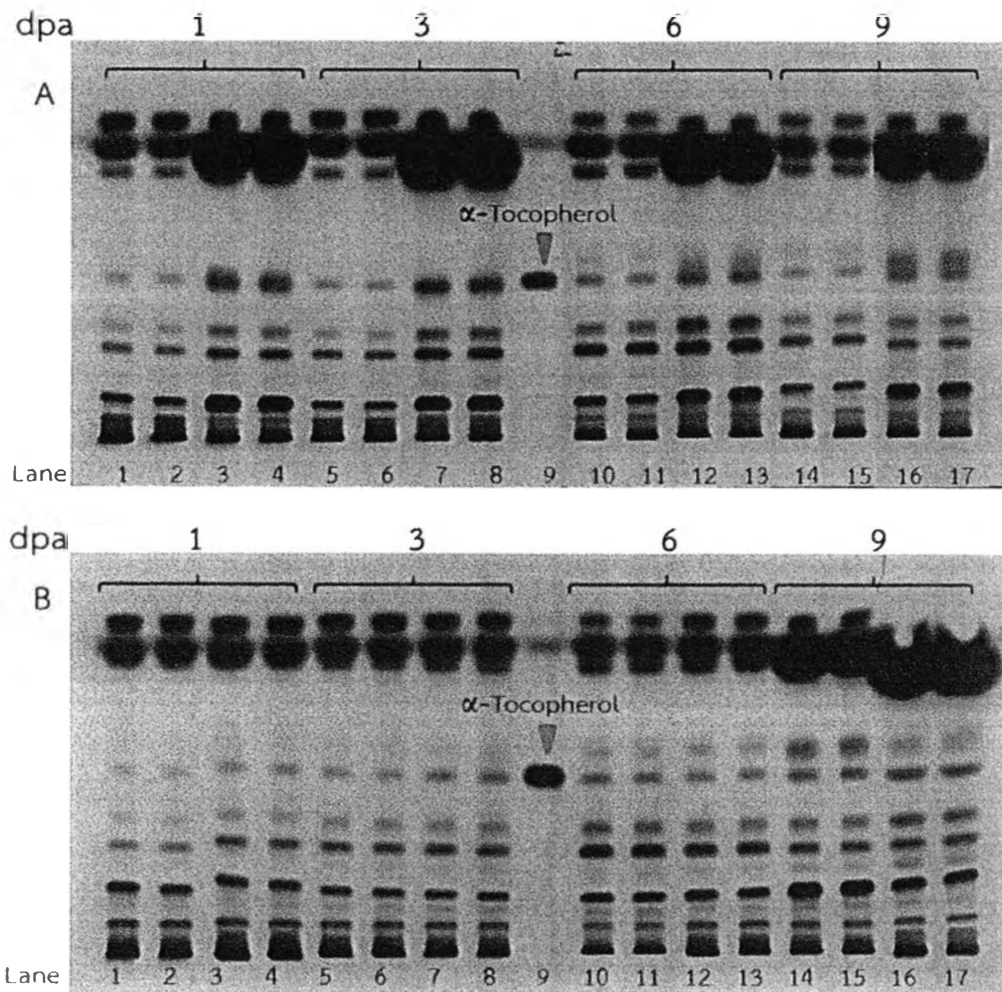


Figure 34 TLC patterns of tomato leaves expressing *alrc2* and *ctl* extracts. The agroinfitrated leaves at 1, 3, 6 and 9 dpa expressing *ctl* (A) and *alrc2* (B) were extracted and separated on the TLC plate. Standard α-tocopherol was used to compare with α-tocopherol from the samples. Each samples was spotted in duplicate. Lane 1 – 2: empty vector (pGWB6) at 1 dpa, Lane 3 – 4: recombinant at 1 dpa, Lane 5 – 6: empty vector (pGWB6) at 3 dpa, Lane 7 – 8: recombinant at 3 dpa, Lane 9: α-Tocopherol standard, Lane 10 – 11: empty vector (pGWB6) at 6 dpa, Lane 12 – 13: recombinant at 6 dpa, Lane 14 – 15: empty vector (pGWB6) at 9 dpa and Lane 16 – 17: recombinant at 9 dpa.

In order to confirm the α-tocopherol synthesis as a result of the overexpression of the recombinant protein, the α-tocopherol extracted from the infiltrated leaves was detected by GC-MS method. The GC-MS chromatogram showed the increase of α-tocopherol, fatty acids, and lipids from the transformed leaves when compared with the control (Figure 35). In addition, during transient overexpression of both genes showed high level of phytol (6.074 min), a precursor of phytyldiphosphate (PDP) which is the substrate of enzyme HPT and also palmitic acid (5.534 min), linoleic acid (6.248 min), α- linoleic acid (6.281 min) and stearic acid (6.359 min) (Figure 36A and 36B), was detected at 3 and 6 dpa corresponding to the increase of α-tocopherol content at retention time of 12.953 min and its intermediate accumulation MPBQ and DMPBQ have retention time of 7.923 and 8.023 min, respectively (Figure 36C and 36D) From GC-MS spectra, it is possible to identify α-tocopherol, MPBQ and DMPBQ, intermediates which are not commercially available, as shown in Figure 37A, B and C, respectively. The MS spectrum of MPBQ revealed quinol head group fragment at m/z 281 and 321 combined with phytyl group fragment at m/z 265 while, the DMPBQ MS spectrum showed quinol head group fragment at m/z 281 and 335 combined with phytyl group fragment at m/z 265 and 155.

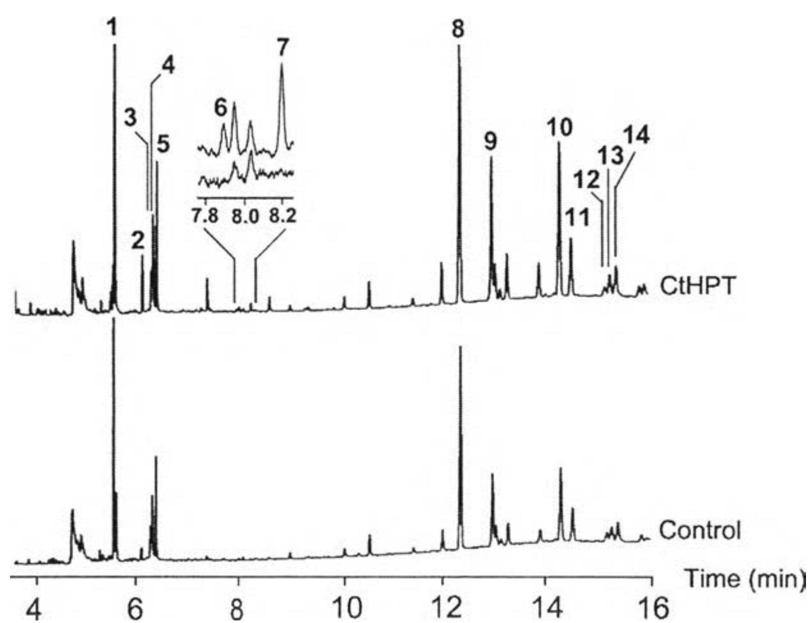
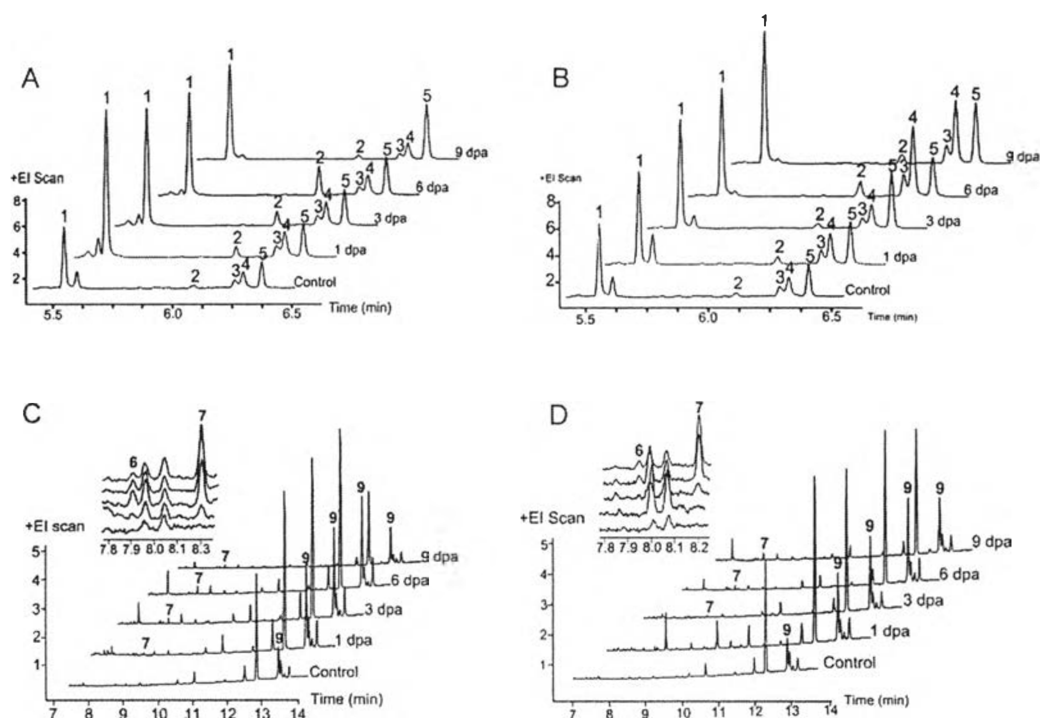


Figure 35 GC-MS chromatogram of infiltrated leaves in 3 dpa of pGWB6::*ctl* showed the increase of metabolites. The numbers on the GC-MS chromatograms are the compounds: 1: palmitic acid (5.534 min), 2: phytol (6.074 min), 3: linoleic acid (6.248 min), 4: α- linoleic acid (6.281 min), 5: stearic acid (6.359 min), 6: MPBQ (7.923 min), 7: DMPBQ (8.203 min), 8: pentacosane (12.321 min), 9: α-tocopherol (12.953 min), 10: nanocosane (14.279 min), 11: α-stigmasterol (14.511 min) 12: β-Sitosterol (15.171 min), 13: Monolinoelaidin (15.258) and 14 :β-Amyrin (15.387 min).

2747256186



2747256186

Figure 36 GC-MS analysis of the chemical profiles of phytol and fatty acids comparing between the transient expression of *ctl* (A) and *alrc2* (B) and α-tocopherol together with intermediates (MPBQ, DMPBQ) involved in the biosynthetic pathway of *ctl* (C) and *alrc2* (D) in tomato leaves at 1, 3, 6 and 9 dpa compare with control (empty vector: pGWB6). The labeled numbers: 1: palmitic acid (5.534 min), 2: phytol (6.074 min), 3: linoleic acid (6.248 min), 4: α- linoleic acid (6.281 min), 5: stearic acid (6.359 min), 6:MPBQ (7.923 min), 7: DMPBQ (8.203 min), 8: pentacosane (12.321 min), 9: α-tocopherol (12.953 min).

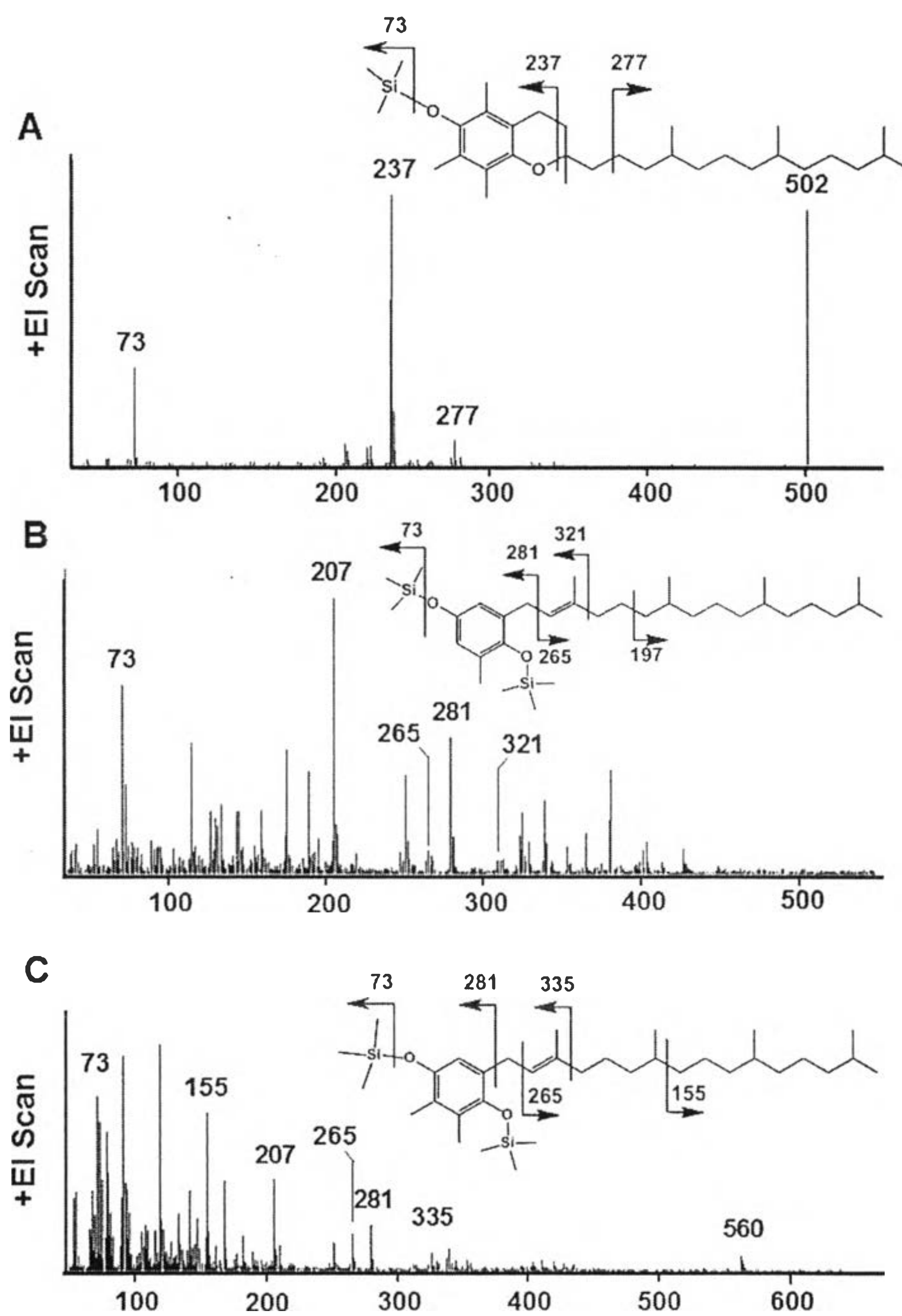


Figure 37 Mass spectra of silylated (A) α-tocopherol (12.955 min), (B) MPBQ (7.923 min) and (C) DMPBQ (8.203 min) from infiltrated leaves.

2747256186